Bayesian Bandits

Reading

- Sutton, Richard S., and Barto, Andrew G. Reinforcement learning: An introduction. MIT press, 2018.
  - http://www.incompleteideas.net/book/the-book-2nd.html
  - Chapter 2
- Slivkins, Aleksandrs. "Introduction to multi-armed bandits." Foundations and Trends in Machine Learning 12.1-2 (2019): 1-286.
  - https://arxiv.org/pdf/1904.07272
  - Chapters 3

Overview

Rensselaer

- In many cases, we might have a prior guess for each action
  - E.g., suppose you have two slightly biased coins
    - You want to determine which one has a higher likelihood of heads
    - Both are probably close to 0.5, so it makes sense to start from 0.5
- In Bayesian methods, we treat the unknown parameter itself as a random variable
- A very different learning paradigm from the alternative where the unknown parameter is treated as a fixed constant
- We'll see how we can use this paradigm in the case of bandits

Bayesian vs. Frequentist Approach

- One of the classical dichotomies in the learning/statistical communities
- Frequentists approach learning problems without any preconceptions and just let the data speak for itself
  - We are trying to learn some parameter (e.g., a coin bias)
    - choose the estimate that best fits the data we have
- Bayesians claim that we should use our prior knowledge about how the world works
  - E.g., a coin is biased but the probability of H is most likely closer to 0.5 than 1
  - Since the prior is not perfect, it is essentially a probability distribution of the parameter value

Bayesian vs. Frequentist Approach, cont'd

- Apart from the philosophical discussion, there are pragmatic considerations as well
- Ultimately, we care about how well algorithms perform on real data
- My advice is not to be too attached to philosophy but pay close attention to what the data is saying
  - If you think your prior is good, but a Bayesian approach doesn't work so well, try to understand why
    - E.g., you used a wrong distribution class, wrong observation model
  - A frequentist approach sounds less biased but it still requires assumptions about your data
    - Linear, sigmoid, etc.
    - Neural networks are the ultimate frequentist tool

Coin Bias Example

- Suppose I want to estimate the probability of a coin being H
- What is the frequentist approach?
  - Flip the coin N times
  - Use the proportion of Hs as your unbiased estimate of the probability of H
    - Bonus points: use Hoeffding's inequality the bound the uncertainty around your estimate

Coin Bias Example, cont'd

- Suppose I want to estimate the probability of a coin being H
- What is the Bayesian approach?
- Model the probability of H as a random variable
  - Denote it by θ
- Suppose I have a prior on θ
  - For simplicity, my prior says θ can only take on 10 values:
$$\mathbb{P}[\theta = 0.5] = p_1, \dots, \mathbb{P}[\theta = 0.6] = p_{10}$$
- Suppose I flip the coin and get a H
  - How do I update my prior?

Probability Aside: Bayes Rule

- Recall the definition of conditional probability:

$$\mathbb{P}[X|Y] = \frac{\mathbb{P}[X,Y]}{\mathbb{P}[Y]}$$

- Can I write $\mathbb{P}[X|Y]$ as a function of $\mathbb{P}[Y|X]$?

$$\mathbb{P}[Y|X] = \frac{\mathbb{P}[X,Y]}{\mathbb{P}[X]}$$

  - i.e.,

$$\mathbb{P}[X,Y] = \mathbb{P}[X]\mathbb{P}[Y|X]$$

- Plugging in the top equation

$$\mathbb{P}[X|Y] = \frac{\mathbb{P}[X]\mathbb{P}[Y|X]}{\mathbb{P}[Y]}$$

- This identity is known as Bayes Rule

Coin Bias Example, cont'd

- For simplicity, my prior says θ can only take on 10 values:

$$\mathbb{P}[\theta = 0.5] = p_1, \dots, \mathbb{P}[\theta = 0.6] = p_{10}$$

- Suppose I flip the coin and get a H
  - How do I update my prior?
    - I want to calculate $\mathbb{P}[\theta = p|R_2 = 1]$ for each $p \in \{0.5, \dots, 0.6\}$
    - Suppose $R_2 = 1$ if I get a H (and 0 otherwise)
- Using Bayes Rule:

$$\mathbb{P}[\theta = p|R_2 = 1] = \\ = \frac{\mathbb{P}[\theta = p]\mathbb{P}[R_2 = 1|\theta = p]}{\mathbb{P}[R_2 = 1]}$$

  - We know $\mathbb{P}[\theta = p]$: it is the prior
  - We know $\mathbb{P}[R_2 = 1|\theta = p] = p$
  - What about $\mathbb{P}[R_2 = 1]$?

Coin Bias Example, cont'd

- Using Bayes Rule:

$$\mathbb{P}[\theta = p | R_2 = 1] =$$

$$= \frac{\mathbb{P}[\theta = p]\mathbb{P}[R_2 = 1 | \theta = p]}{\mathbb{P}[R_2 = 1]}$$

  - We know $\mathbb{P}[\theta = p]$: it is the prior
  - We know $\mathbb{P}[R_2 = 1 | \theta = p] = p$
  - What about $\mathbb{P}[R_2 = 1]$?
  - Using marginalization and conditional probability

$$\mathbb{P}[R_2 = 1] = \sum_p \mathbb{P}[R_2 = 1, \theta = p]$$

$$= \sum_p \mathbb{P}[\theta = p]\mathbb{P}[R_2 = 1 | \theta = p]$$

Coin Bias Example, cont'd

- The final Bayesian update becomes

$$\mathbb{P}[\theta = p | R_2 = 1] = \frac{\mathbb{P}[\theta = p]\mathbb{P}[R_2 = 1 | \theta = p]}{\sum_{p_i} \mathbb{P}[\theta = p_i]\mathbb{P}[R_2 = 1 | \theta = p_i]}$$

- This is known as the posterior distribution of θ
  - Prior → before receiving data
  - Posterior → after receiving data
- What do I do after the next flip?
  - Use the previous posterior as the next prior
- The Bayesian approach thus has a nice iterative implementation

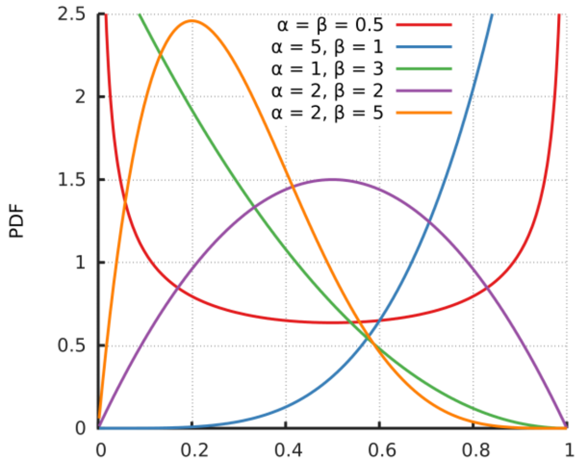
Coin Bias Example, Beta Approach

- What issues do you see with our approach so far?
  - It is constrained to only 10 possibilities for θ
  - I cannot estimate it with higher precision
- Ideally, I will use a continuous distribution so that all real values of θ are possible
  - Let's try the Beta distribution

Probability Aside: The Beta Distribution

Rensselaer

- The Beta distribution models a random parameter that defines the probability of an event (e.g., coin toss)



- It has parameters $\alpha, \beta > 0$, which appear as exponents of the variable and its complement, respectively

- The Beta probability density function (pdf) is

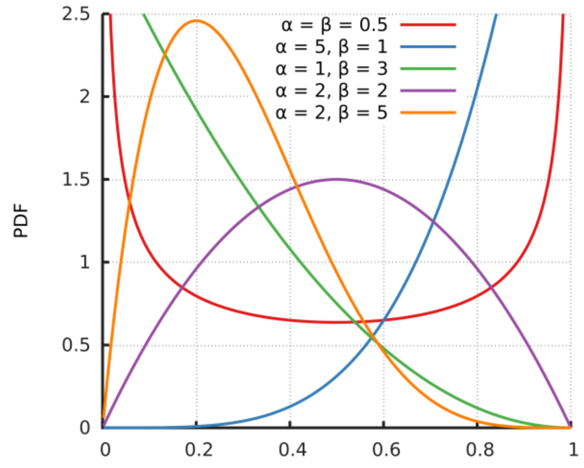
$$p(x; \alpha, \beta) = const \; * x^{\alpha-1}(1-x)^{\beta-1}$$

  – A pdf is almost like a standard probability function
    - Not a probability function since the probability of a single point is 0
    - It's similar to a probability function since it has to integrate to 1

$$\int_{-\infty}^{\infty} p(x; \alpha, \beta) dx = 1$$

Probability Aside: The Beta Distribution

Rensselaer



- The Beta distribution models a random parameter that defines the probability of an event
- It has parameters $\alpha, \beta > 0$, which appear as exponents of the variable and its complement, respectively
- The Beta probability density function (pdf) is

$$p(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\, du}$$

  – Note that the mean is as follows

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x p(x; \alpha, \beta) dx = \frac{1}{1+\beta/\alpha}$$

  - Notation $p(x; \alpha, \beta)$ just makes it explicit what the parameters are

Coin Bias Example, Beta Approach, cont'd

- Suppose my prior for θ is a Beta distribution with parameters α_0, β_0
- Suppose I flip a H as before
- It turns out that Bayes Rule applies to pdfs as well

$$p[x; \alpha_0, \beta_0 | R_2 = 1] =$$

$$= \frac{p[x; \alpha_0, \beta_0] \mathbb{P}[R_2 = 1 | \theta = x]}{\mathbb{P}[R_2 = 1]}$$

$$= const * x^{\alpha_0 - 1}(1 - x)^{\beta_0 - 1} \cdot x$$

$$= const * \ x^{(\alpha_0 + 1) - 1}(1 - x)^{\beta_0 - 1}$$

– where $const = \frac{1}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du \cdot \mathbb{P}[R_2=1]}$

Coin Bias Example, Beta Approach, cont'd

- It turns out that Bayes Rule applies to pdfs as well

$$p[x;\alpha_0,\beta_0|R_2=1] = const * \; x^{(\alpha_0+1)-1}(1-x)^{\beta_0-1}$$

  - where $const = \frac{1}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du \cdot \mathbb{P}[R_2=1]}$

- This is another Beta distribution!
  - with parameters $\alpha_1 = \alpha_0 + 1, \beta_1 = \beta_0$
  - You should make sure $const$ can be simplified to the normalizing constant for the new Beta distribution

- We say the Beta distribution is a conjugate prior for the Bernoulli distribution
  - The posterior and the prior remain in the same probability class, with different parameters

Bayesian Bandits

- Suppose now I have 2 coins and would like to learn which one is more likely to come out as H
  - Can we map this to a bandit problem?
  - Suppose I get a reward of 1 for each H and 0 otherwise
  - Which action brings me a higher reward in expectation?
- In the Bayesian world, each coin's probability of success is a random variable
  - E.g., the probability of coin 1 being H is denoted by θ_1
- Suppose I have a prior on each θ_i
  - For simplicity, my prior says θ_i can only take on 10 values:
$$\mathbb{P}[\theta_i = 0.5] = p_{i,1}, \dots, \mathbb{P}[\theta_i = 0.6] = p_{i,10}$$
- Which coin do you flip next?

Bayesian Bandits, cont'd

- Suppose I have a prior on each θ_i
  - For simplicity, my prior says θ_i can only take on 10 values:
    $$\mathbb{P}[\theta_i = 0.5] = p_{i,1}, \dots, \mathbb{P}[\theta_i = 0.6] = p_{i,10}$$
- Which coin do you flip next?
  - Need to calculate which coin is more likely to flip H
    $\mathbb{P}[\theta_1 \geq \theta_2] =$
    $$= \sum_{p_1 > p_2} \mathbb{P}\left[\theta_1 = p_1, \theta_2 = p_2\right]$$
- If $\mathbb{P}[\theta_1 \geq \theta_2] > 0.5$, then flip coin 1, else coin 2
- Suppose I flip coin 1 and get a reward of 1
  - How do I update θ_1?
  - I want to calculate $\mathbb{P}[\theta_1 = p | R_2 = 1, A_1 = 1]$ for each p

Bayesian Bandits, cont'd

- Using Bayes Rule (same derivation as the 1-coin case):

$$\mathbb{P}[\theta_1 = p | R_2 = 1, A_1 = 1] = $$
$$= \frac{\mathbb{P}[\theta_1 = p]\mathbb{P}[R_2 = 1, A_1 = 1 | \theta_1 = p]}{\mathbb{P}[R_2 = 1, A_1 = 1]}$$

- We know $\mathbb{P}[\theta_1 = p]$
- What about $\mathbb{P}[R_2 = 1, A_1 = 1 | \theta_1 = p]$?
  - Using the definition of conditional probability

$$\mathbb{P}[R_2 = 1, A_1 = 1 | \theta_1 = p] = $$
$$= \mathbb{P}[R_2 = 1 | A_1 = 1, \theta_1 = p]\mathbb{P}[A_1 = 1 | \theta_1 = p]$$

  - We know $\mathbb{P}[R_2 = 1 | A_1 = 1, \theta_1 = p] = p$
  - Also, note that A_1 does not depend on θ_1
    - The action depends only on observed data
    - So $\mathbb{P}[A_1 = 1 | \theta_1 = p] = \mathbb{P}[A_1 = 1]$
    - Finally, $\mathbb{P}[R_2 = 1, A_1 = 1 | \theta_1 = p] = p\mathbb{P}[A_1 = 1]$

Bayesian Bandits, cont'd

- Using Bayes Rule (same derivation as the 1-coin case):

$$\mathbb{P}[\theta_1 = p | R_2 = 1, A_1 = 1] = \\ = \frac{\mathbb{P}[\theta_1 = p]\mathbb{P}[R_2 = 1, A_1 = 1 | \theta_1 = p]}{\mathbb{P}[R_2 = 1, A_1 = 1]}$$

- We know $\mathbb{P}[\theta_1 = p]$ and $\mathbb{P}[R_2 = 1, A_1 = 1|\theta_1 = p]$
- What about $\mathbb{P}[R_2 = 1, A_1 = 1]$?
  - Using marginalization and conditional probability

$$\mathbb{P}[R_2 = 1, A_1 = 1] = \sum_p \mathbb{P}[R_1 = 1, A_1 = 1, \theta_1 = p]$$

$$= \sum_p \mathbb{P}[\theta_1 = p]\mathbb{P}[R_2 = 1, A_1 = 1|\theta_1 = p]$$

$$= \mathbb{P}[A_1 = 1] \sum_p \mathbb{P}[\theta_1 = p]p$$

Bayesian Bandits, cont'd

- So the final Bayesian update is

$$\mathbb{P}[\theta_1 = p | R_2 = 1, A_1 = 1] =$$

$$= \frac{\mathbb{P}[\theta_1 = p]p\mathbb{P}[A_1 = 1]}{\mathbb{P}[A_1 = 1]\sum_{p_i}\mathbb{P}[\theta_1 = p_i]p_i}$$

$$= \frac{\mathbb{P}[\theta_1 = p]p}{\sum_{p_i}\mathbb{P}[\theta_1 = p_i]p_i}$$

  - So the posterior is independent of the algorithm!
    - As soon as we flip coin 1, we perform a standard Bayesian update
    - Regardless of how many times we flipped other coins in between the coin 1 flips
  - Need to calculate for all $p \in \{0.5, \ldots, 0.6\}$

Thompson Sampling

- What challenges do you see with the Bayesian approach?
- Calculating the posterior is not trivial when θ is not finite
  - The posterior distribution may be hard to represent mathematically
    - Assuming a beta prior is one way to resolve this, but it may not always be the right prior
- Calculating the probability $\mathbb{P}[\theta_1 > \theta_2]$ may not even be possible in closed form
  - May require heavy computation to approximate, especially if you have more actions
- The Thompson sampling algorithm addresses/alleviates these challenges

Thompson Sampling, cont'd

- Calculating the probability $\mathbb{P}[\theta_1 > \theta_2]$ may not even be possible in closed form
  - Suppose we know the distribution of each θ_i, call it $\mathcal{D}_{\theta_i}$, but don't have a closed-form expression for $\mathbb{P}[\theta_1 > \theta_2]$
  - We can sample $t_i \sim \mathcal{D}_{\theta_i}$ and then take action corresponding to the largest sampled t_i
- The posterior distribution may be hard to represent mathematically
  - Some distributions have closed-form posteriors, e.g., Gaussian and Beta distributions
    - Often good approximations of many real-life scenarios

Thompson Sampling, cont'd

- Algorithm summary:
  - Start with prior distribution for each θ_i, call it $\mathcal{D}_{\theta_i}$
  - Sample $t_i \sim \mathcal{D}_{\theta_i}$ for each i
  - Take action $a_t = a_{i^*}$,where $i^* = arg\max_i t_i$
  - Observe reward r_{t+1}
  - Update $\mathcal{D}_{\theta_{i^*}}$ using Bayes rule
    - E.g., assuming a Beta prior